MEMORANDUM
RM-5728-PR
APRIL 1969

A MONTE CARLO STUDY OF THE REGRESSION MODEL WITH AUTOCORRELATED DISTURBANCES

Clifford G. Hildreth and John Y. Lu

PREPARED FOR:

UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

53

MEMORANDUM

RM-5728-PR

APRIL 1969

A MONTE CARLO STUDY OF THE REGRESSION MODEL WITH AUTOCORRELATED DISTURBANCES

Clifford G. Hildreth and John Y. Lu

This research is supported by the United States Air Force under Project RAND—Contract No. F44620-67-C-0045—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation

Published by The RAND Corporation

PREFACE

This Memorandum is part of RAND's continuing program to develop basic analytical techniques for application to Air Force problems.

The validity of many economic relations derived by applying the ordinary linear least-squares regression method to time series is often questionable, because the implicit assumption of serially independent disturbances cannot be justified The principal alternative model considered assumes that the disturbances are generated by a first-order autoregressive process. Several estimators under the latter assumption have been suggested, but little is known about their small sample properties. This study describes the relative performance of the estimators based on results of a Monte Carlo experiment.

The Memorandum is intended for operational and economic analysts who deal with time series data. It is assumed that the reader is familiar with basic econometric literature on time series analysis Two potential areas of Air Force application are manpower prediction and demand prediction for spares.

Clifford Hildreth is a consultant to the Logistics Department of RAND.

SUMMARY

Economists interested in analyzing time series have long recognized autocorrelated disturbances as one of the principal hazards that may cause serious inefficiencies in their analyses. In the past decade, several econometricians have studied a statistical model in which the disturbances are assumed to be generated by a simple, first-order autoregressive process, and have proposed several estimators of the unknown parameters. Because little is known of the probability laws governing the estimators and therefore of their relative desirability in various circumstances, and because determining the laws analytically poses severe problems, a study of the behavior of alternative estimators applied to artificially generated data with known parameter values was undertaken, and its results are reported herein.

To generate artificial data for this experiment, eight structures were specified. Each structure differs from the others in one or more of the following aspects: the pattern of observed values of the independent variables (these are arranged in a matrix denoted by Z); the value of the autocorrelation coefficient (ρ); and the sample size. Samples of size 30 were drawn for four structures and samples of size 100 were drawn for four others. For each structure, 300 samples of the selected size were drawn and estimates of unknown parameters were calculated for each sample by five different methods. They were Maximum likelihood (ML) estimators, Theil-Nagar (TN) estimators, approximate Bayes (AB) estimators, Durbin (D) estimators and Least squares (LS) estimators.

Analyses of the performance of the above five estimators leads to several general observations:

1. When ρ is nonnegative, ML, TN and D estimators all have a persistent tendency to underestimate ρ on the average.
2. Judging by the absolute deviations of sample means from their respective true values, the TN estimator of ρ looks slightly better for samples of size 30 and relatively small ρ, i.e., $|\rho| \leq 0.3$; however, ML appears to perform better for samples of size 100 and relatively large absolute values of ρ. The Durbin procedure appears a little less biased than TN for samples with 100 observations, but, in general, it appears least favorable among the three estimators.
3. The TN estimator of ρ has a smaller variance than the ML estimator for samples with 30 observations and relatively small ρ. However, the variance of the ML estimator is smaller for samples of 100 observations and relatively large ρ.
4. On the average, the D estimator of ρ has a larger variance than the other two estimators.
5. The sample means of all the estimators of γ's are similar and are close to their true values, even for samples with as few as 30 observations.
6. Judging by low mean square error, TN estimates of γ's are a little better than ML for samples with 30 observations, but for samples with 100 observations, both estimators perform about the same. The D estimator is slightly worse than both ML and TN estimators regardless of sample size.
7. For samples with only 30 observations and ρ as large as 0.3, the other three estimators do not have advantages over the LS estimator. LS also estimates coefficients well when the columns of Z are "smooth."

Besides examining the performance of various estimators, we also checked the behavior of several commonly used tests of independence of regression disturbances. The tests considered were the Von Neumann ratio test, the Durbin-Watson test, the Theil-Nagar test, the likelihood ratio test, and the test based on the asymptotic distribution of the ML estimate of ρ (we shall call this the $\hat{\rho}$ test). Some general observations are as follows:

8. There were many inconclusive applications of the DW test, as previously noted by both theorists and practical workers.

9. For the sample sizes used in this study, the TN test amounts to rejecting the null hypothesis in those cases where the DW test either rejects or is inconclusive. Inspection of the TN and DW tables reveals that this will be virtually true except for quite small samples.

10. TN rejected a true null hypothesis much too frequently for samples of size 30.

11. The tendency noted above for ML to underestimate ρ was reflected in low frequencies of rejection of true null hypotheses by one-tailed $\hat{\rho}$ tests and high frequencies for two-tailed tests. Thus, the $\hat{\rho}$ test cannot be recommended when based on the asymptotic distribution. In considering this bias in the actual significance level, however, the rejection rates for false hypotheses were relatively large. This suggests that a powerful test can be based on $\hat{\rho}$ if a good approximation to its finite sample distribution can be found.

Hildreth [13] has shown that the ML estimators are asymptotically normal and that the vector, $\hat{\gamma}$, of estimates of coefficients is asymptotically independent of $\hat{\rho}$, $\hat{v}$, the estimators of the autocorrelation coefficient and the variance. It was conjectured that, for many purposes, the asymptotic distribution of $\hat{\gamma}$ would prove a tolerable approximation in the sample sizes often encountered in econometric studies, but that for $\hat{\rho}$ and $\hat{v}$ the asymptotic distributions would be less satisfactory. This tends to be confirmed by the χ^2 goodness-of-fit statistics computed from the generated data.

In conclusion, the reader must be aware that the above observations are descriptive statements of how certain statistics behaved in this particular experiment. Since 300 samples were drawn for each structure, we hope that the observed characteristics are generally representative of these structures. The characteristics of the various structures were chosen to represent a variety of circumstances that might reasonably be encountered in practical work. To know just how representative the structures are, however, would require a careful

survey of applications, and this has not been undertaken. It is desirable that hints furnished by a study such as this be supplanted by analytical results whenever possible. For important properties that remain intractable after further theoretical analysis, additional Monte Carlo experiments are in order.

ACKNOWLEDGMENTS

We wish to express our appreciation to R. J. Clasen, R. H. Mavall and Colleen Giller for programming this Monte Carlo study, and to A. J. Gross and G. S Fishman, who read an earlier draft and offered many useful comments.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS ix

Section
I. INTRODUCTION 1

II. DESIGN OF THE STUDY 3
Maximum Likelihood (ML) 6
Theil-Nagar (TN) 6
Approximate Bayes (AB) 7
Durbin (D) 9

III. RESULTS 10
Performance of the Estimators 10
Summary Tables 10
Comparison of the Various Estimators of ρ 15
Comparison of the Various Estimators of γ's 17
Tests of Significance 20
Approximate Distributions 24

Appendix
A. METHOD FOR GENERATING ARTIFICIAL DATA 31
B. SPECIFICATIONS AND PROPERTIES OF $z_{t,k}$'s 33
C. MAXIMIZING THE LIKELIHOOD FUNCTION 35
D. COMPUTER PROGRAM FOR GENERATING A SAMPLE FOR THE MONTE CARLO STUDY 37

REFERENCES 41

I. INTRODUCTION

Economists have long been concerned that nonindependent disturbances may be a frequent cause of inefficiency in estimates of regression coefficients for time series.[+] In the past decade, several econometricians have studied an alternative model in which the disturbances are assumed to be generated by a simple, first-order autoregressive process, and have proposed several estimators of the unknown parameters.

Because little is known of the probability laws of the estimators and therefore of their relative desirability in various circumstances, and because determining the laws analytically poses severe problems, we undertook a study of the behavior of alternative estimators applied to artificially generated data with known parameter values. Such studies, of course, furnish hints rather than conclusions about the behavior of various statistics. The investigator determines certain structures in advance and generates samples by drawing random components according to a specified probability law, with the aid of tables of random numbers or other random devices. The results may be misleading because of special features of the structures chosen or because statistical accidents occur in generating samples [18, especially pp. 3-5].

The hints from a particular study can be strengthened by drawing many samples for each structure (thus insuring a low probability of misleading statistical accidents), and by examining a wide array of representative structures. Of course, each tactic increases the resources needed, and the study's final design is always a compromise

---

[+] See [3, 6, 7, 17, 21].

between the cost of resources and the desire to make the results as reliable as possible.

In this study, eight structures were chosen, with samples of size 30 drawn for four structures and of size 100 for four others. For each structure, 300 samples of the selected size were drawn, estimates of unknown parameters were calculated for each sample by alternative methods, and characteristics of the resulting frequency distributions of estimates were calculated and tabulated.

Section II completes a sketch of the study's design and gives reasons for some of the choices. Section III presents and discusses the study's results. Appendices A through D describe in some detail the methods used to generate artificial data and to obtain the maximum likelihood estimator.

II. DESIGN OF THE STUDY

The model employed specifies that an observed vector y of order equal to the sample size (30 or 100) comes from a multivariate normal population with mean vector $Z\gamma$ and variance matrix νA,

where Z : a known matrix of order T x K representing T observed values of each of the K independent variables;

γ : a vector of K unknown coefficients to be estimated;

A : a T x T matrix with typical element $a_{st} = [1/(1-\rho^2)]\rho^{|t-s|}$;

ρ : a constant, $|\rho| < 1$, called the autocorrelation coefficient;

ν : a positive constant.

The interpretation is that an element y_t of y is determined as a linear combination of corresponding elements of Z plus a disturbance that is linearly related to the disturbance of the preceding observation; i.e.,

(1) $$y_t = \sum_{k=1}^{K} z_{tk}\gamma_k + u_t ,$$ where

(2) $$u_t = \rho u_{t-1} + v_t , \qquad t = 2, 3, \ldots, T,$$

$$u_1 = \frac{v_1}{\sqrt{1 - \rho^2}} ,$$ and

the v_t are normal, identical, and independent with mean 0 and variance ν.

We chose a sample size of 30 for four structures because many studies of economic time series involve 20 to 40 observations. We chose 100 as the other sample size because autocorrelation is very likely to be present in quarterly or monthly data, and in these cases

the sample size may be much larger--100 or more is not uncommon. And it seemed desirable to have two sample sizes far enough apart so that we might note any tendencies for asymptotic properties to be more nearly realized in the larger samples.

Past theoretical studies [2, 4, 6] show that properties of some suggested procedures depend critically on the value of ρ and on the pattern of Z. It therefore seemed useful to arrange a set of structures that included various combinations of values of ρ and patterns of Z.

In the present model, the principal aspect of Z (other than sample variances of its rows and sample correlations among rows, which are important in any regression situation) that proved important is the relation of its columns to the characteristic vectors of an approximation to the inverse of the variance matrix A [4, pp. 13-18].

If the columns of Z are linear combinations of K characteristic vectors of this modified inverse, then least-squares estimates of γ are best unbiased and tests of $\rho = 0$ (like those of Durbin and Watson) based essentially on a Von Neumann ratio formed from least-squares residuals are uniformly most powerful against alternatives in the interval (0, 1). Furthermore, the characteristic vectors are harmonic series, and if the K characteristic vectors that approximate Z are of low frequency, then one of the approximations employed by Theil and Nagar [19] can be shown to be close.

For these reasons the Z's employed in three of our structures have been formed so that the last three columns (the first column consists entirely of ones in all of our structures) would be approximately equal (see Appendix B for details) to three characteristic vectors

of relatively low frequency, thus insuring that the above conditions approximately hold. These Z's are described as "smooth" (S). For three other structures, called "rough" (R), the Z-matrices are constructed so that they cannot be closely approximated by any K of the characteristic vectors. For the remaining two structures, called "empirical" (E), three rows of Z are taken from observed time series of important economic variables.

The characteristics of our structures cited so far are summarized in Table 1.

Table 1

CHARACTERISTICS OF STRUCTURES

Structure Number	ρ	Nature of Z	Sample Size
1	.3	S	30
2	0	S	30
3	-.7	S	100
4	.7	R	30
5	.3	R	100
6	0	R	100
7	.5	E	30
8	.9	E	100

In all of the structures,

$$\gamma = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } \nu = 1 .$$

Simple sample correlations between columns of Z other than the first vary from -0.349 to 0.937, and sample variances of these columns vary from 0.46 to 0.75. These arrangements insure that the random term contributes substantially to the variation in the dependent variable

in all structures, while letting other structural characteristics vary. See Appendix B for a more detailed account.

For comparison with each other and with ordinary least-squares estimates of parameters, the following estimators were employed.

MAXIMUM LIKELIHOOD (ML)

For the model defined in (1) and (2), the likelihood function is proportional to

$$\phi(\gamma, \rho, \nu) = \nu^{-t/2}(1 - \rho^2)^{1/2} \exp\left[-\frac{1}{2\nu}(y - Z\gamma)' A^{-1}(y - Z\gamma)\right] .$$

Hildreth and Lu [10] suggest one algorithm for maximizing the logarithm of the above function. Computations were originally performed partly by graphs and partly by hand calculations, but a program for digital computers has subsequently been prepared.[†] The authors [10] showed that ML estimates are consistent, and Hildreth [12, 13] subsequently showed that they are asymptotically normal and asymptotically efficient. Klein [14] suggests another algorithm, and Fuller and Martin [8] develop an approximate procedure. It has also been claimed that an iterative procedure suggested by Cochrane and Orcutt [3] converges to ML estimates. The Hildreth-Lu algorithm was used in this study because it contains some safeguards against undetected multiple maxima.

THEIL-NAGAR (TN)

These authors suggest a two-step procedure for estimating the parameters of the model in (1) and (2). Based on an extension of a

[†]The algorithm is described in Appendix D.

procedure they suggest for testing the hypothesis of serial independence [19], they obtain an estimate of the first-order autocorrelation coefficient, say $\tilde{\rho}$. Other parameters are then estimated by applying the classical least-squares regression of

$$(y_t - \tilde{\rho} y_{t-1}) \text{ on } (z_{t1} - \tilde{\rho} z_{t-1,1}), \ldots, (z_{tK} - \tilde{\rho} z_{t-1,K}) .$$

Their procedure for estimating ρ is based on an approximate distribution of the Von Neumann ratio obtained by fitting a β-distribution to approximate moments, after which $\tilde{\rho}$ is obtained by linear interpolation:

$$\tilde{\rho} = \frac{T^2[1 - (1/2)R] + K^2}{T^2 - K^2} ,$$

where R is the Von Neumann ratio defined on p. 31.

Since some approximation errors do not disappear with increasing sample size, the estimator is not consistent. Thus, for any given structure, there must be a sample size for which a consistent procedure (e.g., ML above or D below) becomes superior. Although Theil and Nagar are unable to evaluate all the approximations they employ, their rationalization is generally cogent and it seems important to obtain whatever clues our data contain about the relative performance of this estimator with typical sample sizes.

APPROXIMATE BAYES (AB)

It would have been desirable to compare other estimators with the mean of the Bayesian posterior distribution corresponding to a diffuse prior. Unfortunately, this would have extended the computing task

beyond what could be contemplated in the present study. Instead, the mean of an approximate prior suggested by Zellner and Tiao [22] is used.

From (1) and (2)

$$(3)\quad y_t = y_{t-1}\rho + \sum_{k=1}^{K} z_{tk}\gamma_k - \sum_{k=1}^{K} z_{t-1,k}\gamma_k\rho + v_t \qquad t = 2, 3 \ldots, T .$$

Each of the nonlinear terms $\gamma_k\rho$ in (3) is expanded about the ML estimators, say $\hat{\gamma}_k$ and $\hat{\rho}$, as follows:

$$(4)\qquad \gamma_k\rho \doteq \hat{\gamma}_k\hat{\rho} + (\rho - \hat{\rho})\hat{\gamma}_k + (\gamma_k - \hat{\gamma}_k)\hat{\rho} ,$$

where "$\doteq$" may be read "is approximated by." Inserting (4) into (3) and collecting terms with the same unknown parameters γ_k and ρ, yields

$$(5)\quad y_t - \hat{\rho}\sum_k \hat{\gamma}_k z_{t-1,k} \doteq \rho(y_{t-1} - \sum_k \hat{\gamma}_k z_{t-1,k})$$

$$+ \sum_k \gamma_k(z_{tk} - \hat{\rho} z_{t-1,k}) + v_t ,$$

which is linear in γ_k and ρ. The fitted least squares regression of

$$(y_t - \hat{\rho}\sum_k \hat{\gamma}_k z_{t-1,k}) \text{ on } (y_{t-1} - \sum_k \hat{\gamma} z_{t-1,k}),\ (z_{t1} - \hat{\rho} z_{t-1,1}),\ \ldots,$$

$$(z_{tK} - \hat{\rho} z_{t-1,K})$$

gives estimates of the coefficients ρ and γ's. Since this estimator is an adjustment of the ML estimator, its relation to the latter is of particular interest.

DURBIN (D)

Durbin [5] suggests another two-step procedure. Let y'_t and z'_{tK} be deviations from the respective sample means of y_t and z_{tK}. The procedure involves taking the linear regression of y'_t on y'_{t-1}, z'_{t1}, ..., z'_{tK}, $z'_{t-1,1}$, ..., $z'_{t-1,K}$. The resulting regression coefficient of y'_{t-1} is its estimate of ρ, say $\tilde{\rho}$. We then apply once more the least-squares regression of $(y_t - \tilde{\rho} y_{t-1})$ on $(z_{t1} - \tilde{\rho} z_{t-1,1})$, ..., $(z_{tK} - \tilde{\rho} z_{t-1,K})$. Although these estimators are consistent and asymptotically equivalent to ML, there is room for doubt about their finite sample properties. Equation (3) differs from a standard linear model in having a lagged dependent explanatory variable. It is also clear that the variables on the right will be nearly multicollinear in many economic applications.

III. RESULTS

PERFORMANCE OF THE ESTIMATORS

Summary Tables

Tables 2 and 3 summarize the principal calculations. For each structure and each estimation procedure, the tables show the mean, variance, and mean square error of the 300 estimates for each of the six parameters. Each entry in a column headed "Mean" is the simple arithmetic mean of the 300 estimates of the parameter indicated by the row label and the structure indicated by the row group, using the estimation method indicated by the column group. The "MSE" (mean square error) and the "Var" (variance) columns are similarly set up. For example, the calculation of the entry 0.049 in row three, column five of Table 2 may be indicated

$$(6) \qquad r^{(2)}_{\tilde{\gamma}_3} = \frac{1}{300} \sum_{n=1}^{300} \left(\tilde{\gamma}^{(2)}_{3_n} - M^{(2)}_{\tilde{\gamma}_3}\right)^2 ,$$

where $r^{(2)}_{\tilde{\gamma}_3}$ = the calculated variance of the 300 Theil-Nagar estimates of γ_3 in structure 2,

$\tilde{\gamma}^{(2)}_{3_n}$ = the Theil-Nagar estimate of γ_3 in the n^{th} sample generated by structure 2, and

$M^{(2)}_{\tilde{\gamma}_3}$ the arithmetic mean, 1.008, of these estimators (it appears in row three, column 4).

The corresponding mean square error, 0.049, in column six may be indicated

$$(7) \qquad e^{(2)}_{\tilde{\gamma}_3} = \frac{1}{300} \sum_{n=1}^{300} \left(\tilde{\gamma}^{(2)}_{3_n} - \gamma^{(2)}_3\right)^2 ,$$

Table 2

SUMMARY OF ESTIMATES FOR SAMPLE SIZE 30

Structure			Estimation Methods														
			ML			TN			LS			AB			D		
ID No.	Parameter Estimated	True Value	Mean	Var	MSE	Mean	Var	MSE	Mean	Var	MSE	Mean	Var	MSE	Mean	Var	MSE
2	γ_1	0	-0.015	0.033	0.033	-0.015	0.033	0.033	-0.015	0.033	0.033	-0.013	0.035	0.036	0.018	0.035	0.036
	γ_2	1	0.991	0.077	0.077	0.991	0.076	0.076	0.993	0.071	0.071	0.996	0.084	0.084	0.987	0.076	0.076
	γ_3	1	1.008	0.05[illegible]	0.[illegible]50	1.008	0.049	0.049	1.007	0.049	0.049	1.005	0.054	0.054	1.004	0.052	0.052
	γ_4	1	1.018	0.078	0.078	1.018	0.077	0.078	1.015	0.075	0.075	1.015	0.080	0.080	1.013	0.078	0.078
	ρ	0	-0.120	0.039	0.053	-0.062	0.034	0.038	-	-	-	-0.121	0.039	0.053	-0.136	0.037	0.055
	[illegible]	1	0.834	0.051	0.079	0.838	0.052	0.078	0.876	0.[illegible]2	0.067	1.041	0.077	0.079	0.806	0.050	0.088
	[illegible]	1	1.001	0.073	0.073	1.006	0.075	0.075	1.011	0.069	0.069	1.249	0.111	0.173	0.967	0.072	0.073
1	γ_1	0	-0.008	0.065	0.065	-0.007	0.065	0.065	-0.0[illegible]	0.065	0.065	-0.002	0.073	0.073	-0.011	0.070	0.070
	γ_2	1	0.995	0.116	0.117	0.996	0.114	0.114	0.996	0.113	0.113	0.998	0.130	0.130	0.990	0.123	0.123
	γ_3	1	0.989	0.060	0.060	0.989	0.060	0.060	0.988	0.061	0.061	0.988	0.069	0.069	0.989	0.066	0.066
	γ_4	1	1.013	0.102	0.102	1.0[illegible]	0.102	0.102	1.013	0.102	0.102	1.009	0.117	0.112	1.012	0.113	0.113
	ρ	0.3	0.153	0.041	0.062	0.1[illegible]6	0.035	0.046	-	-	-	0.156	0.041	0.062	1.109	0.042	0.078
	[illegible]	1	0.835	0.053	0.080	-0.837	0.053	0.080	[illegible]992	0.070	0.082	1.040	0.085	0.087	0.803	0.052	0.091
	[illegible]	1	1.002	0.076	0.076	1.004	0.076	0.076	1.029	0.093	0.094	1.248	0.122	0.184	0.963	0.075	0.076
3	γ_1	0	0.002	0.130	0.130	0.001	0.129	0.129	0.002	0.129	0.129	-0.001	0.142	0.142	0.000	0.133	0.133
	γ_2	1	1.080	0.491	0.497	1.079	0.480	0.487	1.086	0.580	0.588	1.087	0.505	0.513	1.082	0.506	0.512
	γ_3	1	0.931	0.386	0.391	0.929	0.373	0.378	0.938	0.459	0.463	0.934	0.395	0.399	0.962	0.395	0.401
	γ_4	1	1.004	0.201	0.201	1.010	0.189	0.189	0.996	0.226	0.226	[illegible]992	0.293	0.293	1.004	0.213	0.213
	ρ	0.5	0.298	0.042	0.083	0.309	0.032	0.068	-	-	-	[illegible]308	0.044	0.081	0.268	0.047	0.101
	[illegible]	1	0.797	0.054	0.095	0.799	0.055	0.095	0.920	0.105	0.112	0.972	0.083	0.083	0.758	0.049	0.108
	[illegible]	1	0.936	0.078	0.080	0.939	0.079	0.081	1.062	0.140	0.144	1.190	0.120	0.156	0.909	0.070	0.079
4	γ_1	0	0.016	0.375	0.376	0.020	0.359	0.359	0.[illegible]0	0.367	0.367	0.048	0.372	0.374	0.033	0.437	0.438
	γ_2	1	0.999	0.036	0.036	0.996	0.041	0.041	0.985	0.123	0.124	0.999	0.037	0.037	1.004	0.043	0.043
	γ_3	1	0.999	0.033	0.033	0.997	0.037	0.037	0.970	0.099	0.100	0.997	0.033	0.033	1.004	0.036	0.036
	γ_4	1	1.005	0.026	0.026	1.007	0.027	0.027	1.036	0.077	0.079	1.004	0.026	0.026	1.012	0.029	0.029
	ρ	0.7	0.513	0.032	0.040	0.512	0.026	0.062	-	-	-	0.606	0.033	0.042	0.552	0.036	0.058
	[illegible]	1	0.843	0.062	0.087	0.870	0.068	0.085	1.391	0.358	0.511	1.029	0.095	0.095	0.792	0.058	0.101
	[illegible]	1	1.012	0.089	0.089	1.044	0.098	0.100	1.605	0.477	0.843	1.235	0.137	0.192	0.850	0.083	0.086

Table 3

SUMMARY OF ESTIMATES FOR SAMPLE SIZE 100

Structure			Estimation Methods														
			ML			TN			LS			AB			D		
ID No.	Parameter Estimated	True Value	Mean	Var	MSE	Mean	Var	MSE	Mean	Var	MSE	Mean	Var	MSE	Mean	Var	MSE
3	γ_1	0	0.000	0.004	0.004	0.000	0.004	0.004	0.001	0.004	0.004	0.000	0.004	0.004	0.000	0.004	0.004
	γ_2	1	0.997	0.006	0.007	0.997	0.006	0.006	0.997	0.010	0.010	0.998	0.007	0.007	0.996	0.007	0.007
	γ_3	1	1.006	0.007	0.007	1.007	0.007	0.007	1.010	0.009	0.009	1.006	0.007	0.007	1.006	0.007	0.007
	γ_4	1	1.004	0.006	0.006	1.004	0.006	0.006	1.003	0.007	0.007	1.004	0.006	0.006	1.003	0.006	0.006
	c	-0.7	-0.696	0.006	0.006	-0.672	0.006	0.007	-	-	-	-0.696	0.006	0.006	-0.694	0.006	0.006
	[illegible]	1	0.953	0.023	0.026	0.956	0.024	0.026	1.926	0.279	1.137	1.008	0.026	0.026	0.934	0.022	0.026
	[illegible]	1	1.003	0.025	0.025	1.006	0.027	0.027	2.006	0.303	1.315	1.061	0.029	0.033	0.983	0.024	0.025
6	γ_1	0	-0.011	0.009	0.009	-0.011	0.009	0.009	-0.011	0.009	0.009	-0.012	0.009	0.010	-0.011	0.009	0.009
	γ_2	1	1.001	0.012	0.012	1.001	0.012	0.012	1.002	0.012	0.012	1.001	0.012	0.012	1.003	0.014	0.014
	γ_3	1	1.005	0.013	0.013	1.005	0.013	0.013	1.006	0.013	0.013	1.005	0.013	0.013	1.007	0.015	0.016
	γ_4	1	1.004	0.014	0.014	1.004	0.014	0.014	1.001	0.014	0.014	1.004	0.015	0.015	1.005	0.015	0.016
	c	0	-0.039	0.010	0.011	-0.026	0.009	0.010	-	-	-	-0.039	0.010	0.011	-0.034	0.009	0.011
	[illegible]	1	0.945	0.018	0.021	0.945	0.018	0.021	0.955	0.018	0.020	1.004	0.021	0.021	0.935	0.018	0.022
	[illegible]	1	0.995	0.020	0.020	0.995	0.020	0.020	0.995	0.020	0.020	1.057	0.023	0.026	0.985	0.019	0.020
5	γ_1	0	0.010	0.018	0.016	0.010	0.018	0.018	0.009	0.018	0.018	0.008	0.018	0.018	0.007	0.019	0.019
	γ_2	1	0.992	0.020	0.020	0.993	0.020	0.020	0.995	0.023	0.023	0.991	0.020	0.020	0.986	0.024	0.024
	γ_3	1	0.996	0.020	0.020	0.996	0.020	0.020	0.996	0.021	0.021	0.999	0.020	0.020	0.990	0.022	0.022
	γ_4	1	1.009	0.020	0.020	1.009	0.019	0.019	1.012	0.022	0.022	1.010	0.019	0.020	1.003	0.022	0.022
	c	0.3	0.270	0.012	0.013	0.263	0.010	0.012	-	-	-	0.270	0.012	0.013	0.264	0.011	0.012
	[illegible]	1	0.960	0.017	0.018	0.961	0.017	0.018	1.047	0.025	0.027	1.022	0.019	0.020	0.949	0.017	0.019
	[illegible]	1	1.011	0.019	0.019	1.012	0.019	0.019	1.091	0.027	0.035	1.076	0.021	0.027	0.999	0.019	0.019
8	γ_1	0	0.031	1.551	1.552	0.056	1.106	1.109	0.055	1.032	1.035	-0.079	6.456	6.462	-0.108	9.128	9.140
	γ_2	1	1.017	0.192	0.192	1.015	0.270	0.271	1.076	0.996	1.002	1.017	0.193	0.193	1.017	0.191	0.191
	γ_3	1	0.996	0.100	0.100	0.998	0.129	0.129	0.931	1.024	1.028	0.994	0.100	0.101	0.996	0.098	0.098
	γ_4	1	0.996	0.052	0.052	1.003	0.062	0.062	1.077	0.441	0.447	0.995	0.052	0.052	0.998	0.055	0.055
	c	0.9	0.870	0.005	0.006	0.797	0.005	0.016	-	-	-	0.871	0.006	0.036	0.822	0.005	0.011
	[illegible]	1	0.987	0.024	0.025	1.020	0.031	0.031	3.584	2.717	9.40[illegible]	1.032	0.027	0.028	0.944	0.021	0.024
	[illegible]	1	1.039	0.027	0.028	1.074	0.034	0.040	3.735	2.948	10.431	1.086	0.030	0.037	0.994	0.023	0.023

where $\gamma_3^{(2)}$ = the true value of γ_3, 1, in structure 2.

Two sets of estimates of the variance, ν, were obtained for each structure-method combination. The first is the quotient of the sum of squares of residuals over the number of observations. Empirical means, variances and mean square errors for estimates calculated in this way appear in the upper rows labeled ν in each of the four sections of the tables.

For methods other than LS, the second set of estimates of ν (figures in parentheses) are calculated by dividing each sum of squares of residuals by T-5 instead of T. Fitting 5 parameters to achieve a low sum of squares tends to make the resulting sum less than that which would correspond to true values of ρ and γ. Since the estimates are nonlinear, one does not know that this is the appropriate adjustment, but it seems a reasonable one to try.

For LS, the second set of estimates of ν are the sum of squares of residuals divided by T-4. This is what someone who applied LS would ordinarily use to estimate the variance. For $\rho \neq 0$ it is known to be biased, but the bias could not be computed without knowing the true value of ρ.

Table 2 includes the structures involving 30 observations, and Table 3 includes those with 100 observations in each sample. The structures in each table are arranged in order of increasing value of ρ.

Part of the information about relative MSEs in Tables 2 and 3 is presented more simply in Table 4. The first column corresponding to each method contains MSE averages for estimates of the four γ's for each structure and for various sets of structures. For instance, 0.059

Table 4

AVERAGES OF MEAN SQUARE ERROR

Classification of Structures		ρ	ML γ's	ML ρ	ML Avg	TN γ's	TN ρ	TN Avg	LS γ's	AB γ's	AB ρ	AB Avg	D γ's	D ρ	D Avg
T=30	2	0	0.059	0.053	0.058	0.059	0.038	0.055	0.057	0.064	0.053	0.062	0.060	0.055	0.059
	1	0.3	0.086	0.062	0.081	0.085	0.046	0.077	0.085	0.096	0.062	0.089	0.093	0.078	0.090
	7	0.5	0.305	0.083	0.260	0.296	0.068	0.250	0.351	0.336	0.081	0.285	0.314	0.101	0.271
	4	0.7	0.118	0.040	0.102	0.116	0.062	0.105	0.167	0.168	0.042	0.143	0.137	0.058	0.121
	Avg		0.142	0.060	0.126	0.139	0.054	0.122	0.165	0.166	0.060	0.145	0.151	0.073	0.135
T=100	3	-0.7	0.006	0.006	0.006	0.006	0.007	0.006	0.008	0.006	0.006	0.006	0.006	0.006	0.006
	6	0	0.012	0.011	0.012	0.012	0.010	0.012	0.012	0.013	0.011	0.013	0.014	0.011	0.013
	5	0.3	0.019	0.013	0.018	0.019	0.012	0.018	0.021	0.020	0.013	0.019	0.022	0.012	0.020
	8	0.9	0.474	0.006	0.380	0.393	0.016	0.318	0.878	1.702	0.006	1.362	2.371	0.011	1.899
	Avg		0.128	0.009	0.104	0.108	0.011	0.089	0.230	0.435	0.009	0.350	0.603	0.010	0.485
All structures			0.135	0.035	0.115	0.124	0.033	0.106	0.198	0.301	0.035	0.248	0.377	0.042	0.310
Smooth (1, 2, 3)			0.050	0.040	0.048	0.050	0.030	0.046	0.050	0.055	0.040	0.052	0.053	0.046	0.052
Rough (4, 5, 6)			0.050	0.021	0.044	0.049	0.028	0.045	0.067	0.067	0.022	0.058	0.058	0.027	0.051
Empirical (7, 8)			0.390	0.045	0.320	0.345	0.042	0.284	0.615	1.019	0.044	0.824	1.343	0.056	1.085
$\|\rho\| \geq 0.5$ (3, 4, 7, 8)			0.226	0.034	0.187	0.203	0.038	0.170	0.351	0.553	0.034	0.449	0.707	0.044	0.574
$\|\rho\| < 0.5$ (1, 2, 5, 6)			0.044	0.035	0.042	0.044	0.027	0.041	0.044	0.048	0.035	0.046	0.047	0.039	0.046

is the average of four mean square errors of γ's estimated by the ML method for structure 2, and 0.142 is the average of these mean square errors over all structures with 30 observations in each sample. For methods other than LS the second column contains MSEs of estimates of ρ for each structure and averages for selected groups of structures. Each entry in the third column is a weighted average of the corresponding entries in the first two columns and represents the average MSE for estimates of all five coefficients for the indicated method and structure (or group of structures). Mean square errors for estimates of ν are excluded in Table 4 since they depend on adjustments for fitted coefficients, and the appropriate adjustments for our nonlinear estimates are not known.

Comparison of the Various Estimators of ρ

To compare the biases of the various estimators, the pertinent Monte Carlo information was extracted from Tables 2 and 3 and summarized in Table 5. Inspection of the table leads to several general observations.

Table 5

MEANS OF DIFFERENT ESTIMATORS[a] OF ρ

Structure		ρ	ML	TN	D
T=30	2	0	-0.120	-0.062	-0.136
	1	0.3	0.153	0.194	0.109
	7	0.5	0.298	0.309	0.268
	4	0.7	0.613	0.512	0.552
T=100	3	-0.7	[illegible]	-0.672	-0.694
	6	0	-0.039	-0.026	-0.034
	5	0.3	0.270	0.263	0.264
	8	0.9	0.870	0.797	0.822

[a]The AB estimator was excluded because it is an adjustment of the ML procedure and appears to be numerically close to the ML estimators for most samples.

A. *When ρ is nonnegative, all three estimators persistently tend to underestimate ρ on the average. Since we have only one structure with negative ρ and the three estimators come close to the true value, we have negligible evidence for this case.*

B. *Judging by the absolute deviations of sample means from their respective true values, the TN estimator of ρ looks slightly better for samples with 30 observations and relatively small ρ, i.e., $|\rho| < 0.3$; however, ML appears to perform better for samples with 100 observations and relatively large absolute values of ρ. The Durbin procedure appears to be a little better than TN for samples with 100 observations, but overall it appears least favorable among the three estimators.*

Since an estimator's performance depends not only on the magnitude of its average bias but also on its variance, we have computed the ratios of the mean square errors of TN and D estimates over those of the ML estimates. The results, presented in Table 6, tend to confirm the above observations regarding the relative performances of the three estimators.

Table 6

RATIOS OF MEAN SQUARE ERRORS OF TN AND D ESTIMATORS OF ρ TO ML ESTIMATOR

	Structure	ρ	MSE of TN estimate of ρ / MSE of ML estimate of ρ	MSE of D estimate of ρ / MSE of ML estimate of ρ
T=30	2	0	0.717	1.038
	1	0.3	0.742	1.258
	7	0.5	0.819	1.217
	4	0.7	1.550	1.450
T=100	3	-0.7	1.167	1.000
	6	0	0.909	1.000
	5	0.3	0.923	0.923
	8	0.9	2.667	1.833

C. *The MSE of the TN estimator of ρ is smaller than that of the ML estimator for samples with relatively small ρ and with 30 observations. The ML estimator, however, has a smaller MSE for samples with relatively large ρ and 100 observations. Referring back to Tables 2 and 3, one sees that TN variances are consistently smaller than ML variance for sample size 30 and slightly smaller in two cases with sample size 100. The generally smaller MSEs for ML estimates with 100 observations are therefore due to smaller biases.*

D. *The D estimator seems inferior to the other two estimators in terms of the mean square error ratios.*

<u>Comparison of the Various Estimators of γ's</u>

E. *The sample means of all the estimators of γ's are similar and are close to their true values. The LS estimator is known to be unbiased. The other estimators also seem to show very little bias even with a 30-observation sample.*

To examine the relative efficiency of the various γ estimators, we divided the average mean square error of γ's for each of the three estimators, ML, TN and D, by that of the LS estimator. This gives some indication of what an investigator will gain if he uses one of the more complicated methods instead of the ordinary least squares method. Results are presented in Table 7. For instance, the first entry in the table, 1.035, was obtained by dividing 0.059 by 0.057. These mean square average errors for γ's are given in Table 4.

Table 7

RELATIVE EFFICIENCY OF DIFFERENT ESTIMATORS OF γ's COMPARED TO LS ESTIMATOR

Structure		ρ	ML/LS	TN/LS	D/LS
T=30	2	0	1.035	1.035	1.053
	1	0.3	1.012	1.000	1.094
	7	0.5	0.869	0.843	0.895
	4	0.7	0.707	0.695	0.820
T=100	3	-0.7	0.750	0.750	0.750
	6	0	1.000	1.000	1.167
	5	0.3	0.905	0.905	1.048
	8	0.9	0.540	0.448	2.700

F. *Judging by the relative efficiency, TN is a little better than ML for samples of size 30 and about the same for samples of size 100. The D estimator performs slightly worse than both the ML and TN estimators.*

G. *For samples with only 30 observations and relatively small value of ρ (0.3), the other three estimators do not have advantages over the LS estimator.*

Interpretation of comparisons for groups of structures is complicated because the number of structures that could be investigated prevented construction of a balanced design reflecting all of the properties considered important. Thus the three structures with smooth Z's include two with T = 30 and one with T = 100, while the three with rough Z's include two larger samples and one smaller one. Comparison of the smooth and rough rows, therefore, indicates only that the effect of smoothness in Table 4 is small relative to sample size in our experiment.

A little better hint can be obtained by averaging MSEs for structures 1 and 3 and comparing these with averages for 4 and 5; the results are shown in Table 8.

Table 8

AVERAGE MSE FOR ALL COEFFICIENTS AND SELECTED STRUCTURES

Structure Combination	ML	TN	LS	AB	D
1, 3 (S)	0.043	0.042	0.047	0.048	0.048
4, 5 (R)	0.060	0.061	0.094	0.081	0.071

The comparisons in Table 8 are a little more meaningful than the S and R rows of Table 4, since each row of Table 8 refers to a pair of structures with T values 30 and 100 and $|\rho|$ equal to 3 and 7 (see the description of structures in Table 1). This shows a tendency for lower MSE with smooth independent variables, particularly for LS. It is clear, however, that any conclusions on effect of smoothness based on data from the present study would be very tenuous. This should be investigated further analytically and, if necessary, by Monte Carlo trials specifically designed for this purpose.

AB estimates are obtained by adjusting ML estimates. From Tables 2 through 4, it appears that, on the average, the adjustment worsens the estimates, at least when judged by MSE. It is also of interest to know whether or not the adjustment is typically large or small. That it was less than 0.05 in most cases in the present study is indicated by Table 9, which contains frequencies of the differences in AB and ML estimates of ρ and of γ_2.

Table 9

COMPARISON OF MAXIMUM LIKELIHOOD ESTIMATES WITH APPROXIMATE BAYES ESTIMATES

Differences of 2 Estimators	Structures							
	T=30				T=100			
	2	1	7	4	3	6	5	8
ρ(ML) - ρ(AB)								
-0.50								
-0.50 ~ -0.20								
-0.20 ~ -0.05	3	5	7	2				2
-0.05 ~ 0.05	296	286	289	256	300	300	300	295
0.05 ~ 0.20	1	9	4	41				3
0.20 ~ 0.50				1				
0.50 ~								
γ_2(ML) - γ_2(AB)								
-0.50			9					
-0.50 ~ -0.20	5	6	37					
-0.20 ~ -0.05	70	79	63					9
-0.05 ~ 0.05	158	135	80	300	300	300	300	283
0.05 ~ 0.20	63	76	67					8
0.20 ~ 0.50	4	4	39					
0.50 ~			5					
True values of ρ	0	0.3	0.5	0.7	-0.7	0	0.3	0.9

It should be noted that the above observations and others to follow are, in the first instance, descriptive statements of how certain

statistics behaved in this one experiment. Since 300 samples were drawn for each structure, we hope that the observed characteristics are generally representative of these structures. It is unknown how well these structures represent those commonly encountered in practice and how many of the properties we have noted will hold for different structures. Thus it is desirable that hints furnished by these studies be supplanted by precise analytical results as quickly and completely as possible. For important properties that remain intractable, further Monte Carlo experiments with different structures are in order.

ML, TN and D seem to understate ρ systematically (at least for nonnegative ρ), suggesting that a systematic adjustment in each estimator might improve its accuracy, especially for small samples. This seems worth pursuing, but the authors believe that further analysis of the distributions of the two estimators is in order before recommendations are formulated. For the ML estimator, the matter is discussed a little further in connection with the discussion on tests of goodness of fit.

The tendency for maximum likelihood to give better estimates than alternative procedures when $|\rho|$ is large is confirmed by a study conducted independently by David F. Reilly [16].

TESTS OF SIGNIFICANCE

Although this study's emphasis is on estimator performance, it would have been wasteful not to have used the data generated to check the behavior of commonly used tests of significance as well. Accordingly, Tables 10 and 11 present the fraction of samples that, for each of several tests, rejects the null hypothesis $\rho \leq 0$ (one-sided) or

Table 10

MONTE CARLO APPROXIMATION TO PROBABILITIES THAT VARIOUS TESTS (ONE-TAILED) REJECT THE NULL HYPOTHESIS $\rho \leq 0$ OR $\rho = 0$

Structure	ρ	Sample Size = 30					Sample Size = 100				
		VN Test	TN Test	DW Test[a]		Large Sample $\hat{\rho}$	VN Test	TN Test	DW Test[a]		Large Sample $\hat{\rho}$
Intended Significance Level--1 percent											
3	-0.7	--	--	--	--	--	0	0	0	(0)	0
2&6[b]	0	0.008	0.024	0.003	(0.021)	0.003	0.008	0.010	0.003	(0.007)	0.007
1&5[c]	0.3	0.130	0.260	0.027	(0.234)	0.087	0.664	0.766	0.503	(0.263)	0.666
7	0.5	0.294	0.524	0.107	(0.418)	0.283	--	--	--	--	--
4	0.7	0.767	0.890	0.506	(0.384)	0.843	--	--	--	--	--
8	0.9	--	--	--	--	--	1.000	1.000	1.000	(0.000)	1.000
Intended Significance Level-- 5 percent											
3	-0.7	--	--	--	--	--	0	0	0	(0)	0
2&6[b]	0	0.030	0.077	0.007	(0.070)	0.017	0.020	0.044	0.010	(0.034)	0.017
1&5[c]	0.3	0.307	0.494	0.123	(0.371)	0.220	0.860	0.888	0.800	(0.087)	0.850
7	0.5	0.558	0.737	0.270	(0.467)	0.523	--	--	--	--	--
4	0.7	0.924	0.957	0.737	(0.220)	0.937	--	--	--	--	--
8	0.9	--	--	--	--	--	1.000	1.000	1.000	(0.000)	1.000

[a] The numbers in parenthesis beside the DW results are the fraction of cases in which this test was inconclusive. The two-tailed DW test was performed at the 2-percent level.

[b] Structure 2 has samples of size 30 and structure 6 has samples of size 100.

[c] Structure 1 has samples of size 30 and structure 5 has samples of size 100.

Table 11

MONTE CARLO APPROXIMATION TO PROBABILITIES THAT VARIOUS TESTS (TWO-TAILED) REJECT THE NULL HYPOTHESIS $\rho = 0$

Structure	ρ	Sample Size = 30				Sample Size = 100			
		VN Test	DW Test[a]	Large Sample $\hat{\rho}$	Likelihood Ratio	VN Test	DW Test[a]	Large Sample $\hat{\rho}$	Likelihood Ratio
		Intended Significance Level--1 percent							
3	-0.7	--	-- --	--	--	1.000	1.000 (0.000)	1.000	1.000
2&6[b]	0	0.010	0.007 (0.134)	0.030	0.018	0.007	0.003 (0.034)	0.014	0.008
1&5[c]	0.3	0.090	0.027 (0.238)	0.060	0.028	0.564	0.503 (0.263)	0.554	0.450
7	0.5	0.218	0.107 (0.418)	0.204	0.127	--	-- --	--	--
4	0.7	0.684	0.506 (0.384)	0.818	0.704	--	-- --	--	--
8	0.9	--	-- --	--	--	1.000	1.000 (0.000)	1.000	1.000
		Intended Significance Level--5 percent							
3	0.7	--	-- --	--	--	1.000	1.000 (0.000)	1.000	1.000
2&6[b]	0	0.050	0.017 (0.260)	0.134	0.054	0.033	0.020 (0.077)	0.073	0.033
1&5[c]	0.3	0.217	0.067 (0.351)	0.160	0.097	0.820	0.663 (0.184)	0.910	0.664
7	0.5	0.450	0.203 (0.427)	0.408	0.257	--	-- --	--	--
4	0.7	0.867	0.637 (0.307)	0.917	0.820	--	-- --	--	--
8	0.9	--	-- --	--	--	1.000	1.000 (1.000)	1.000	1.000

[a]The numbers in parenthesis beside the DW results are the fraction of cases in which this test was inconclusive. The two-tailed DW test was performed at the 2-percent level.

[b]Structure 2 has samples of size 30 and structure 6 has samples of size 100.

[c]Structure 1 has samples of size 30 and structure 5 has samples of size 100.

$\rho = 0$ (two-sided) for each structure. Each entry is the fraction of 300 samples in which the indicated test rejected the null hypothesis in question. The result of the Durbin-Watson test is sometimes inconclusive (see [6], p. 409). The proportion of cases in which this occurred is indicated in parentheses beside the entry indicating the proportion in which the null hypothesis was rejected.

The Von Neumann ratio test [20], the Theil-Nagar test [19], and the Durbin-Watson test [6, 7] have frequently been used in econometrics. All are based on the Von Neumann ratio of mean successive difference to sample variance. An investigator using likelihood methods would find $\hat{\rho}$, the ML estimate of ρ, or the likelihood ratio a natural test statistic.[+]

Hildreth shows [12] that $\hat{\rho}$ is asymptotically normally distributed with mean ρ and variance $\frac{1-\rho^2}{T}$. Hence, a test based on this asymptotic distribution may be applied to these null hypotheses by referring to a normal distribution with zero mean and variance $1/T$.

The likelihood ratio test has been applied by assuming that $-2 \log \lambda$ (where λ is the likelihood ratio) is approximately χ_1^2. The likelihood ratio is, of course, only useful for two-tailed tests.

Since the Durbin and Watson tables do not provide for a 1-percent two-tailed test, the results shown for the two-tailed DW test are for an intended 2-percent significance level. Theil and Nagar did not recommend that two-tailed tests be performed using their table; but, because their tabulated critical points are almost identical to the critical points d_u in the Durbin-Watson tables (see result D below), one could obtain the results for two-tailed TN tests by adding the

[+]The refinement of the Theil-Nagar test suggested by Henshaw [9] came to our attention after computations were under way.

regular entry in the DW column to the parenthetical entry immediately to the right.

Principal results indicated by Table 8 are the following:

A. *There were many inconclusive applications of DW, as previously noted by both theorists and practical workers.*

B. *The low empirical significance levels associated with one-tailed δ tests when ρ is actually zero, and the high levels for two-tailed tests, are consistent with the tendency previously noted for δ to be negative when $\rho = 0$. This suggests that tests based on δ cannot be recommended for moderate-sized samples until a better approximation to its distribution is developed.*

C. *For samples of size 30, the tabulated power of the TN test must be discounted because the test rejects a true null hypothesis much more frequently than it should.*

D. *A comparison of the TN and DW columns for one-tailed tests indicates that the proportion rejected by TN is equal (within rounding error) to the proportion rejected by DW plus the proportion inconclusive by DW. Inspection of their tables indicates that the TN critical values are within 0.01 of the corresponding upper DW critical values except for samples smaller than 20. Thus, in practice, applying TN is virtually the same as applying DW and rejecting the null hypothesis if the DW procedure either indicates rejection or is inconclusive.*

E. *The LR test based on the asymptotic distribution is not very powerful for samples of size [illegible]0 and, for samples of size 100, the rejection rate for true hypotheses is lower than the intended significance level.*

APPROXIMATE DISTRIBUTIONS

As mentioned in Sec. I, Hildreth [13] has shown that the ML estimators are asymptotically distributed according to a multivariate normal law with ρ, δ, θ mutually asymptotically independent. The asymptotic variances are[+]

[+] Limits of these moments are known to equal the corresponding moments of the limiting distribution, since it can be shown that fourth moments of the ML estimators are bounded. See [13], p. 10.

$$\lim_{T\to\infty} E\sqrt{T}\,(\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)' \quad \nu V \tag{8}$$

$$\lim_{T\to\infty} E\sqrt{T}\,(\hat{\rho} - \rho)^2 = (1 - \rho^2)$$

$$\lim_{T\to\infty} E\sqrt{T}\,(\hat{\nu} - \nu)^2 = 2\nu^2 ,$$

where γ, ρ, ν are the ML estimates and

$$V = \left(\lim_{T\to\infty} \frac{1}{T} Z'A^{-1}Z\right)^{-1} .$$

It was conjectured that, for many purposes, the asymptotic distribution of $\hat{\gamma}$ would prove a tolerable approximation in samples of the size often encountered in econometric studies, but that for $\hat{\rho}$ and $\hat{\nu}$ the asymptotic distributions would be less satisfactory.

The χ^2 goodness-of-fit statistics listed in Tables 12, 13, and 14 tend to confirm this conjecture. Table 12 was constructed by determining 13 intervals for each component of γ and computing the expected frequency of estimates in each interval under the assumption that the estimator was distributed according to its asymptotic law. Adjacent intervals with small expected frequencies were combined to follow Cochrane's recommendation that no more than 20 percent of the remaining intervals should have expected frequencies smaller than 5. This determined the "df" entries.

Observed frequencies in each interval were then tabulated and a χ^2 value for each estimator was computed by the familiar one-way formula,

$$\chi^2 = \sum_{i=1}^{I} \frac{(E_i - O_i)^2}{E_i} , \tag{9}$$

Table 12

χ^2 STATISTICS FOR ASYMPTOTIC DISTRIBUTION OF $\hat{\gamma}$'s

Structure		$\hat{\gamma}_1$ d.f.	$\hat{\gamma}_1$ 5% Points	$\hat{\gamma}_1$ χ^2	$\hat{\gamma}_2$ d.f.	$\hat{\gamma}_2$ 5% Points	$\hat{\gamma}_2$ χ^2	$\hat{\gamma}_3$ d.f.	$\hat{\gamma}_3$ 5% Points	$\hat{\gamma}_3$ χ^2	$\hat{\gamma}_4$ d.f.	$\hat{\gamma}_4$ 5% Points	$\hat{\gamma}_4$ χ^2
I=30	2	8	15.5	10.9	10	18.3	13.6	8	15.5	2.7	10	18.3	11.7
	1	10	18.3	7.4	10	18.3	3.6	8	15.5	10.7	10	18.3	10.7
	7	10	18.3	3.9	11	19.7	10.0	11	19.7	15.2	9	16.9	3.8
	4	11	19.7	12.9	8	15.5	4.7	8	15.5	4.0	8	15.5	11.4
I=100	3	4	9.5	6.8	6	12.6	2.0	6	12.6	1.1	6	12.6	5.9
	6	6	12.6	7.4	6	12.6	8.6	6	12.6	1.5	6	12.6	8.3
	5	6	12.6	8.8	6	12.6	5.6	6	12.6	6.6	6	12.6	9.5
	8	10	18.3	34.6	9	16.9	7.0	10	18.3	13.2	8	15.5	6.1

where O_i, E_i are respectively the observed and expected frequencies, and the number of intervals is I.

Table 13 was constructed similarly except that alternative theoretical distributions were used to determine expected frequencies in calculating the χ^2 statistics appearing in the last two columns. For the column headed β^*, a modified β- distribution[†] was determined by

[†]Let $f_\beta(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}$ for $0 < x < 1$ be a β-density and let $w = 2x - 1$. Then

$$g(w) = \frac{1}{2^{p+q-1} B(p,q)} (1+w)^{p-1}(1-w)^{q-1} ,$$

for $-1 \le w \le 1$ is the density of w.

$$Ew = \frac{p-q}{p+q}, \qquad Var(w) = \frac{4pq}{(p+q)^2(p+q+1)} .$$

Setting $Ew = \rho$, $Var\ w = \frac{1-\rho^2}{T}$ yields

$$p = \frac{T(1+\rho)}{2} , \qquad q = \frac{T(1-\rho)}{2}$$

Setting $Ew = \rho - \frac{3}{T}(1+\rho)$, $Var\ w = \frac{1-\rho^2}{T} + \frac{9}{T^2}(1+\rho)$ yields

$$p = \frac{1}{2}\left[\frac{A^2BT^2}{(1+\rho)T(1-\rho)+9} + A\right], \qquad q = \frac{1}{2}\left[\frac{AB^2T^2}{(1+\rho)[T(1-\rho)+9]} + B\right] ,$$

with $A = (1 + \rho - \frac{3}{T} - \frac{3\rho}{T})$, $B = (1 - \rho + \frac{3}{T} + \frac{3\rho}{T})$.

transforming the variable so that the interval of nonnegative density was (-1, 1) rather than (0, 1), and then determining the remaining free parameters to make the mean and variance equal to their asymptotic values, ρ and $\frac{1-\rho^2}{T}$.

Table 13

χ^2 STATISTICS FOR $\hat{\rho}$

Structure		ρ	d.f.	5% Points	Calculated Values of χ^2: Asymptotic	β^*	β^{**}
T=30	2	0	8	15.5	196.7	194.23	8.95
	1	0.3	8	15.5	317.9	277.56	5.12
	7	0.5	8	15.5	945.6	686.57	28.28
	4	0.7	6	12.6	70.6	194.69	103.79
T=100	3	-0.7	4	9.49	1.8	4.05	11.40
	6	0	6	12.6	52.3	51.78	10.10
	5	0.3	6	12.6	40.7	37.76	5.08
	8	0.9	4	9.49	131.8	131.89	149.29

The theoretical distribution used in calculating the column headed β^{**} was similar to that for β^* except that the mean was set equal to $\rho - \frac{3(1+\rho)}{T}$ and the variance equal to $\frac{1-\rho^2}{T} + \frac{9}{T^2}(1+\rho)$. The latter expressions crudely approximate the means and variances that appear in Tables 2 and 3 for various values of ρ. The β^* and β^{**} distributions were superficial guesses made in a quick attempt to find a better approximation to the distribution of $\hat{\rho}$ in typical samples. Though β^{**} does reduce the "badness" of fit substantially, except for high values of ρ, it does not look promising to us, and we believe an attempt to determine more properties of the finite-sample distribution of $\hat{\rho}$ analytically should precede further attempts to find a better approximation.

Table 14 also contains χ^2 values calculated from the asymptotic distribution (this time of ϑ) and another which, it was guessed, might provide a better fit for typical samples. The asymptotic distribution is normal with mean 1 and variance 2/T; the alternative was obtained by assuming that $T\hat{\vartheta}$ was χ^2_{T-K-1}. The latter amounts to treating ρ as though it entered linearly. Though this approximation did fit well for samples of size 30, neither it nor the asymptotic distribution was a good approximation for samples of size 100. Here, again, closer study of properties of the actual finite sample distribution is in order.

Table 14

χ^2 STATISTICS FOR $\hat{\vartheta}$

Structure		Asymptotic d.f.	Asymptotic 5% Points	Asymptotic χ^2	Gamma d.f.	Gamma 5% Points	Gamma χ^2
T=30	2	10	18.3	157.48	9	16.9	9.94
	1	10	18.3	162.74	9	16.9	4.12
	7	10	18.3	246.43	9	16.9	19.78
	4	10	18.3	158.66	9	16.9	6.56
T=100	3	6	12.6	41.24	7	14.1	27.97
	6	6	12.6	51.55	7	14.1	168.25
	5	6	12.6	33.84	7	14.1	13.70
	8	6	12.6	13.04	7	14.1	37.90

One reason for examining the fit of approximations to the maximum likelihood estimators is the conjecture that it may be possible to construct a useful approximate Bayesian procedure for applications of this model if a sufficiently simple and accurate approximation can be found.[*] Prospects for such a procedure are enhanced if the esti-

[*] See pp. 426-427 of [11].

mators $\hat{\gamma}$, $\hat{\rho}$, $\hat{\nu}$ are "approximately" independent in samples encountered in practice. The aspect of independence that can most readily be checked (and is quite possibly the most important aspect if utility functions are approximately linear) is linear noncorrelation.

For this prospect, Table 15 is highly encouraging. Simple correlation coefficients between $\hat{\rho}$, $\hat{\nu}$, and components of $\hat{\gamma}$ are presented for each structure.

Table 15

SIMPLE CORRELATION COEFFICIENTS

Pairs of Estimators		Structures							
		1	2	3	4	5	6	7	8
$\hat{\rho}$	$\hat{\gamma}_1$	0.084	-0.004	0.105	-0.049	-0.056	0.003	0.076	-0.076
$\hat{\rho}$	$\hat{\gamma}_2$	-0.096	0.024	0.005	-0.051	0.008	0.070	0.036	-0.072
$\hat{\rho}$	$\hat{\gamma}_3$	-0.047	0.060	0	0.048	0.005	0.045	-0.031	0 008
$\hat{\rho}$	$\hat{\gamma}_4$	-0.045	0.058	0.104	0.084	0.082	-0.003	-0.061	-0.128
$\hat{\nu}$	$\hat{\gamma}_1$	0.078	-0.023	0.007	-0.100	0.037	-0.025	0	-0.156
$\hat{\nu}$	$\hat{\gamma}_2$	-0.028	0.028	0.026	0.073	-0.053	0	0.016	-0.038
$\hat{\nu}$	$\hat{\gamma}_3$	0.109	0.045	-0.043	-0.104	0.053	0 025	0.007	0.001
$\hat{\nu}$	$\hat{\gamma}_4$	-0.045	-0.031	-0.082	-0.027	-0.043	0.087	-0.019	0.022
$\hat{\nu}$	$\hat{\rho}$	0.096	0.177	-0.067	0.186	0.142	-0.001	0.297	0.195

For 300 observations, the significance points of the sampling distribution of simple correlation coefficients under the assumptions of normality and $\rho = 0$ are $\pm$ 0.1133 at the 5-percent level. Consequently, among the 81 correlation coefficients examined, only 7 rejected the null hypothesis. Since 5 out of the 7 rejected cases involve correlation coefficients between $\hat{\rho}$ and $\hat{\nu}$, we probably cannot assume that

they are independent in most samples encountered in practice; however, the elements of $\hat{\gamma}$ are approximately uncorrelated with $\hat{\rho}$ and $\hat{\nu}$.

Appendix A

METHOD FOR GENERATING ARTIFICIAL DATA

Our procedure for generating time series with known properties consisted of the following steps:

a. We first decided on a particular combination of parameter values for the model represented by (1) and (2). All together, eight different combinations of the parameter values were considered (see Table 1, p. 5).

b. The values assumed by all the explanatory variables, Z_{tk}, for $t = 1, \ldots, T$ and $k = 1, \ldots, K$, were also specified. In general, these values were varied from one structure to another.

c. We then generated T random numbers, each of which was normally and independently distributed with mean 0 and variance 1.

d. These random numbers were used as independent disturbances of the model and T observations were obtained on the dependent variables, conditioned on the assumed values of the parameters and the explanatory variables. The y_t's thus generated, together with the corresponding Z_{tk}'s, constituted a sample.

e. The four estimating methods (ML, TN, AB, and D), plus the least squares (LS) method, were each applied to the above sample for estimating γ_k's, ν, ρ and the Von Neumann ratio statistic R.[+]

[+]
$$R = \frac{\sum_{t=2}^{T} \left[\tilde{u}_t - \tilde{u}_{t-1}\right]^2}{\sum_{t=1}^{T} \tilde{u}_t^2}\,,$$
where $\tilde{u}_t$ is the LS estimate of the disturbance u_t.

f. For each of the eight structures, steps (c) through (e) were repeated 300 times. The resulting 300 sets of parameter estimates became the basic data for our sampling experiment with respect to that structure.[1]

Procedures for generating data as described above were programmed in FORTRAN IV. For those interested in further experimentation using samples with different characteristics, usage of the program is described in Appendix D.

---

[1] It took approximately 25 minutes of IBM 7044 computer time to obtain 300 sets of parameter estimates.

Appendix B

SPECIFICATIONS AND PROPERTIES OF $z_{t,k}$'s

For reasons discussed in Sec. II, we constructed three different types of independent variables. The values of the independent variables in structures 1, 2, and 3 are such that they approximate the conditions favorable to TN and LS: those for structures 4, 5, and 6 do not. The independent variables for structures 7 and 8 were based on empirical time series.

To specify the independent variables of structures 1 to 6, let us define a typical element in the jth characteristic vector of an approximation to the inverse of the variance matrix A as [4, p. 17]

$$(11) \qquad R(j, t) = \cos\left[\frac{2t - 1}{T} j\pi\right] \qquad t = 1, 2, \ldots, T .$$

Using the above notation, the independent variables for structures 1 to 6 are presented in Table 16.

Characteristics of the assumed values of the independent variables for various structures are summarized in Table 17. Note that the sample variances of all the Z variables are less than 1, and that their sample correlation coefficients are small except for those based on the empirical data.

Table 16

SPECIFICATIONS OF THE Z MATRICES FOR STRUCTURES 1 TO 6

Structure	Independent Variables			
	1 Constant Term	2^a	3	4
1	$R(0,t)$	$R(2,t)+\epsilon_t$	$R(11,t)+\epsilon_t$	$R(5,t)+\epsilon_t$
2	$R(0,t)$	$R(2,t)+\epsilon_t$	$R(11,t)+\epsilon_t$	$R(5,t)+\epsilon_t$
3	$R(0,t)$	$R(4,t)+\epsilon_t$	$R(23,t)+\epsilon_t$	$R(10,t)+\epsilon_t$
4	$R(0,t)$	$\frac{1}{2}\sum_{j\in J_1} R(j,t)$	$\frac{1}{2}\sum_{j\in J_2} R(j,t)$	$\frac{1}{2}\sum_{j\in J_3} R(j,t)$
		$J_1 = \{1,4,9,15,20,25\}$	$J_2 = \{2,5,11,17,21,26\}$	$J_3 = \{3,6,13,18,23,27\}$
5	$R(0,t)$	$\frac{1}{2}\sum_{j\in J_1} R(j,t)$	$\frac{1}{2}\sum_{j\in J_2} R(j,t)$	$\frac{1}{2}\sum_{j\in J_3} R(j,t)$
		$J_1 = \{1,4,8,15,20,60\}$	$J_2 = \{2,5,11,17,21,72\}$	$J_3 = \{3,6,13,18,23,85\}$
6	Same as Structure 5			

NOTE: A random element was added to each of the independent variables (other than the constant term) in the first three structures. For Structures 7 and 8, the original data were taken from three sets of empirical time series from U.S. Statistical Abstract: (1) wholesale price index; (2) numbers of immigrants; (3) exports of foodstuff. The values of these independent variables had been adjusted so that their sample variances would be 0.75, which is small relative to the variance of the random disturbances. This was intended for easier interpretation of the sampling experiment results.

[a] ϵ_t is a normal deviate with mean 0 and variance $\frac{1}{2}$.

Table 17

CHARACTERISTICS OF Z VARIABLES: MEANS, VARIANCES, AND CORRELATION COEFFICIENTS

Z Variables	Structures						
	1	2	3	4	5&6	7	8
Means							
Z_2	0.066	-0.061	0.006	0.000	0.000	-0.000	0.000
Z_3	0.015	-0.024	-0.007	0.025	0.020	0.000	0.001
Z_4	0.015	-0.024	-0.007	0.025	0.020	0.000	0.000
Variances							
Z_2	0.466	0.539	0.553	0.750	0.750	0.750	0.750
Z_3	0.711	0.616	0.578	0.727	0.750	0.750	0.748
Z_4	0.460	0.477	0.606	0.736	0.750	0.750	0.750
Correlation Coefficients							
Z_2, Z_3	0.113	0.131	0.028	0.010	0.015	0.937	0.798
Z_2, Z_4	0.026	-0.153	-0.040	-0.349	0.012	0.698	0.291
Z_3, Z_4	0.048	0.089	0.089	0.218	-0.025	0.704	0.108

Appendix C

MAXIMIZING THE LIKELIHOOD FUNCTION

The likelihood function of this study differs from the one developed in [10] only in the added assumption that the u_t in Eq. (1) has a stationary distribution.+ In particular, it is assumed that all of the u_t's are normally distributed with zero means and equal variances $[1/(1 - \rho^2)] \omega$.

To obtain ML estimates of γ and ρ the following procedure was used: Assume a particular value of ρ, and compute the corresponding estimates of γ and the sum of squared residuals from (12) and (13), respectively:

$$\text{(12)} \qquad \tilde{\gamma}(\rho) = [Z'A^{-1}Z]^{-1} [Z'A^{-1}y]$$

$$\text{(13)} \qquad \tilde{S}(\rho) = [y - Z\tilde{\gamma}(\rho)]' A^{-1} [y - Z\tilde{\gamma}(\rho)] .$$

This is done for a number of selected values of ρ, and the value which approximately minimizes $\tilde{S}(\rho)$ is determined to a desired accuracy. If the value, $\hat{\rho}$, of ρ that minimizes $\tilde{S}(\rho)$ can be found, then $\hat{\rho}$ and $\hat{\gamma} = \tilde{\gamma}(\hat{\rho})$ are ML estimates of ρ and γ.

A computer program was developed for numerical search of a minimum point of $\tilde{S}(\rho)$ in the interval $-1 < \rho < 1$. Since $\tilde{S}(\rho)$ is a polynomial of high degrees in ρ, a procedure was provided for safeguarding against multiple minima. This was done by examining the successive first-differences of $\tilde{S}(\rho)$ evaluated in the above interval at an increment of

---

+The model based on this assumption of stationarity is discussed in some detail in Append. A of [10] and [13], pp. 2-8.

0.01. If the values of these first differences should change sign more than once, we would conclude that the function probably did not have a unique minimum. It should be noted that no multiple-minima situation appeared to exist in all the samples generated for the study.

After we were reasonably assured of having no multiple minima, numerical search for the minimum sum of squares of residuals was carried out as follows:

(a) $\tilde{S}(\rho)$ is initially evaluated at $\rho = \rho_0^{(1)} - d$, $\rho_0^{(1)}$, $\rho_0^{(1)} + d$, where $d > 0$.

(b) Pick the value of ρ that corresponds to the smallest $\tilde{S}(\rho)$ in Step (a) above. Call this value $\rho_0^{(2)}$.

Steps (a) and (b) make up the first iteration. In the second iteration, $\tilde{S}(\rho)$ is evaluated at $\rho = \rho_0^{(2)} - \frac{d}{2}$, $\rho_0^{(2)}$, $\rho_0^{(2)} + \frac{d}{2}$. In general, in the ith iteration, the function is being evaluated at

$$\rho = \rho_0^{(i)} - \frac{d}{2^{i-1}}, \ \rho_0^{(i)}, \ \rho_0^{(i)} + \frac{d}{2^{i-1}} .$$

In our program, we set

$$\rho_0^{(1)} = 0, \ d = 0.5 \text{ and } i = 1, 2, \ldots, 10.$$

A FORTRAN program for the above search procedure is given in Appendix D.

Appendix D

COMPUTER PROGRAM FOR GENERATING A SAMPLE FOR THE MONTE CARLO STUDY

A computer program[†] was developed for generating the independent variables and dependent variable for a given combination of parameter values. (See Tables 1 and 16.)

This appendix describes how to prepare inputs to this program. The correspondence between the notation used below and that of the main text is as follows:

RHO	=	true value of ρ
GAM	=	vector of true values of γ_k's
KZV[i]	=	j in Eq. (11) for specifying the ith independent variables in structures 1, 2, and 3.
KZV_k[i]	=	j in Eq. (11) for specifying the kth element of the ith independent variable for structures 4, 5, and 6.

INPUT

Five input cards and three methods are explained. The first three cards remain the same and provide input for all three methods. The use of cards 4 and 5 varies, depending on the method.

First Card. Contains 7 integers each of field width 3:

Col. 1-3: K (usually $\leq$ 4; program modification required if greater than 4)

Col. 4-6: T ($\leq$200)

Col. 7-9: NTIMES = N = number of cases to be run.

[†]This was programmed by R. J. Clasen of the Computer Sciences Department at RAND.

Col. 10-12: NRN (>0). This means that NRN-1 runs will be copied from tape NTAP2 onto tape NTAP and then this run will be written as the NRNth run on tape NTAP. If NRN = 1, NTAP2 need not be specified.

Col. 13-15: ITYPE = 1 if first method of Z matrix input is used. It specifies the Z matrices for Structures 1, 2, and 3.

ITYPE > 1 if second method of Z matrix input is used. It specifies the Z matrices for Structures 4, 5, and 6.

ITYPE = 0 if third method of Z matrix input is used. It specifies the Z matrices of Structures 7 and 8.

The value for ITYPE is the number of KZV cards to be read in when ITYPE > 1.

Col. 16-18: NTAP--the FORTRAN tape unit number of the binary tape on which the results will be written.

Col. 19-21: NTAP2--the FORTRAN tape number of the old tape which is copied onto NTAP. Usually NTAP = 8 and NTAP2 = 9, or NTAP2 = 8 and NTAP = 9.

Second Card. Columns 1-12 contain the value of RHO punch with a decimal point. (FORMAT (F12.6)).

Third Card. Contains the vector GAM (=GAMMA), with 12 columns per number (6F12.6).

Method One (ITYPE = 1)

Fourth Card (Method 1). Contains the KZV vector with 3 columns per integer (24I3). The ith column of Z is generated by

$$Z[i] = R\left[KZV[i]\right] + \varepsilon(i) ,$$

$$\text{where } \varepsilon(i) = \begin{cases} 0 & i = 1 \\ \varepsilon_o & i > 1 \end{cases}$$

and ε_o is read in by the next card.

Fifth Card (Method 1). Columns 1-12 contain the value for ε_o (F12.6).

Method 2 (ITYPE > 1)

The input for method 2 consists of the first three cards above, plus ITYPE cards of the form of the fourth card of method 1. Let K be the ith number on the jth such card. Then $KZV_k[i]$,

$$Z[1] = \text{vector of all ones.}$$

$$Z[i] = \frac{1}{2} \sum_{k=1}^{ITYPE} R\left[KZV_k[i]\right] \text{ for } i > 1.$$

Note that each KZV card contains space for K numbers, but that KZV[1] is never used (KZV[2] is in columns 4-6, etc.).

Method 3 (ITYPE = 0)

The Z matrix is read in from T cards, each card containing a row of the Z matrix. The T cards follow the 3 cards that are common to all methods. The first column of Z is set to 1; hence each card need contain only K-1 numbers. FORMAT 203 in the MAIN (LU) program is used. Currently, this format is (10X, 8F10.5), but this may change in the future for the convenience of card punching.

INPUT FOR THE TAPE POSITIONING AND PRINTING PROGRAM

The input for this program consists of one card with three integers punched in fields of three-column width (3I3).

Col. 1-3: NTAP = the FORTRAN unit on which the tape is mounted

Col. 4-6: NRN

Col. 7-9: NOGO

If NOGO = 0, the program will rewind NTAP, and will then space NRN-1 runs forward on the tape, so that the tape will be sitting at the

beginning of the run NRN. If NOGO $\neq$ 0, the tape will be positioned as above, but run NRN will then be printed in a compact form. After printing is completed, the tape will be rewound.

REFERENCES

[1] Aitken, A. C., "On Least Squares and Linear Combinations of Observations," Proceedings of the Royal Society of Edinburgh, Vol. 55, 1934-35, pp. 42-48.

[2] Anderson, T. W., "On the Theory of Testing Serial Correlation," Skandinavisk aktuarietidskrift, Vol. 31, 1948, pp. 88-116.

[3] Cochrane, D., and G. H. Orcutt, "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, Vol. 44, No. 245, March 1949, pp. 32-61.

[4] Chipman, J. S., The Problem of Testing for Serial Correlation in Regression Analysis: The Story of a Dilemma, Technical Report 4, Department of Economics, University of Minnesota, 1965.

[5] Durbin, J., "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, Series B, Vol. 22, No. 1, 1960, pp. 139-153.

[6] -----, and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression. I," Biometrika, Vol. 37, 1950, pp. 409-428.

[7] -----, "Testing for Serial Correlation in Least Squares Regression. II," Biometrika, Vol. 38, 1951, pp. 159-178.

[8] Fuller, W. A., and J. E. Martin, "The Effects of Autocorrelated Errors on the Statistical Estimation of Distributed Lag Models," Journal of Farm Economics, Vol. 43, 1961, pp. 70-82.

[9] Henshaw, R. C., Jr., "Testing Single-Equation Least Squares Regression Models for Autocorrelated Disturbances," Econometrica, Vol. 34, No. 3, July 1966, pp. 646-660.

[10] Hildreth, C. G., and J. Y. Lu, Demand Relations with Autocorrelated Disturbances, Technical Bulletin 276, Agricultural Experiment Station, Michigan State University, 1960.

[11] Hildreth, C. G., "Bayesian Statisticians and Remote Clients," Econometrica, Vol. 31, No. 3, July 1963, pp. 422-438.

[12] -----, Asymptotic Distribution of Maximum Likelihood Estimators in Linear Models with Autoregressive Disturbances, The RAND Corporation, RM-5059-PR, July 1966.

[13] -----, Asymptotic Distribution of Maximum Likelihood Estimators in a Linear Model with Autoregressive Disturbances, Working Paper 3, National Science Foundation Project 0620-5359, University of Minnesota, 1968 (mimeographed). To appear in Annals of Mathematical Statistics.

[14] Klein, L. R., A Textbook of Econometrics, Row, Peterson and Co., Evanston, Illinois, 1953.

[15] Ladd, G. W., Experiments with Autoregressive Error Estimation, Research Bulletin 533, Agricultural and Home Economics Experiment Station, Iowa State University, 1965.

[16] Reilly, David F., Evaluation of the Small Sample Properties of Five Alternative Estimation Methods When the Errors are Correlated, Discussion Paper No. 87, University of Pennsylvania, Department of Economics.

[17] Stone, R., The Measurement of Consumers' Expenditures and Behavior in the United Kingdom, 1920-1938, Cambridge University Press, Cambridge, 1954.

[18] Summers, Robert, "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," Econometrica, Vol. 33, No. 1, January 1965, pp. 1-41.

[19] Theil, H., and A. L. Nagar, "Testing the Independence of Regression Disturbances," Journal of the American Statistical Association, Vol. 56, No. 296, December 1961, pp. 793-806.

[20] Von Neumann, J., "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," Annals of Mathematical Statistics, Vol. 12, 1941, pp. 367-395.

[21] Wold, H., and L. Jureen, Demand Analysis, John Wiley and Sons, New York, 1953.

[22] Zellner, A., and G. C. Tiao, "Bayesian Analysis of the Regression Model with Autocorrelated Errors," Journal of the American Statistical Association, Vol. 59, No. 307, September 1964, pp. 763-778.

DOCUMENT CONTROL DATA

1 ORIGINATING ACTIVITY THE RAND CORPORATION	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP
3. REPORT TITLE A MONTE CARLO STUDY OF THE REGRESSION MODEL WITH AUTOCORRELATED DISTRUBANCES	
4. AUTHOR(S) (Last name, first name, initial) Hildreth, Clifford, G. and John Y. Lu	

5. REPORT DATE April 1969	6a. TOTAL No. OF PAGES 53	6b. No. OF REFS. 22
7. CONTRACT OR GRANT No. F44620-67-C-0045	8. ORIGINATOR'S REPORT No. RM-5728-PR	

9a. AVAILABILITY/LIMITATION NOTICES DDC-1	9b. SPONSORING AGENCY United States Air Force Project RAND

10. ABSTRACT

Description of the relative performance of estimators based on the results of a Monte Carlo experiment, under the assumption that disturbances are generated by a first-order autoregressive process. To generate artifical data for the experiment, eight structures were specified: samples of size 30 were drawn for four structures; samples of size 100 for the other four. For each structure, 300 samples were drawn and estimates of unknown parameters were calculated for each sample by five different methods, namely, maximum likelihood, Theil-Nager, approximate Bayes, Durbin, and least squares estimators. The task was first to examine the performance of the various estimators and second, to check the behavior of several commonly used tests of independence regression analysis. Characteristics of the various structures were chosen to represent a variety of circumstances that might be reasonably encountered in practical work.

11. KEY WORDS

Inventory control
Econometrics
Statistical methods and processes
Monte carlo
Models
Regression analysis